

大体框架

城市区域表征旨在从多源异构数据中提取城市区域的低维向量表示，捕捉空间、语义和功能特征，用于支持城市发展中涌现出的各种下游任务，例如城市功能区域识别、人口预测、交通流量预测、犯罪预测、土地使用分类和GDP预测等。近年来，随着移动设备、社交媒体和地理信息系统的普及，大规模城市数据不断涌现，使得基于数据驱动的城市区域建模成为可能。同时，随着深度学习技术的进步，对于特征的提取方法也越来越多。近年来，随着大语言模型（LLM）的兴起，LLM被整合到城市计算中，用于增强区域表征的语义理解和多模态融合。

区域表征范式

城市区域表征范式总结为以下5个层次：原始数据层、特征提取层、表征学习层、融合层和任务层。

原始数据层

原始数据层为城市区域表征提供底层感知数据基础，主要包括以下几类：

- 物理空间数据：包括遥感影像、街景图片、OSM（建筑形态、路网）等，反映城市的空间结构和视觉特征。**例：**RegionDCL、Urban2Vec
- 社会行为数据：包括GPS轨迹、人类移动、出租车轨迹、交通流、人群热力、社交签到等，反映了城市的动态过程。**例：**MTE、EvolveURE。
- 语义与文本数据：包括POI、社交媒体、新闻描述等，主要提供语义线索。**例：**Urban2Vec
- 复合与衍生数据：基于各类多模态对齐的复合数据，以及从多源数据衍生的时空图结构。**例：**MTGRR
- 环境与外部数据：气象、空气质量、突发事件、政策、节假日等，用于增强时空预测的鲁棒性。**例：**Terra

依据区域间–区域内分类



对每一种数据进行解释？

eg:

POI数据：商店、餐厅、公共设施等位置信息，反映了区域的功能和用途，POI数据常常来自地图服务，可用于描述区域的功能标签。

社交媒体签到与评论：地理标记社交媒体（微博、推特、INS等）产生的签到、评论、和图片等数据，可以反映人群活动强度和偏好。

移动轨迹与GPS数据：来自出租车、共享单车、智能手机等的移动轨迹数据记录了人群的时空行为，是捕捉区域互联模式和人群规律的重要数据源。利用大量轨迹数据可以挖掘出区域间交通流动关系，并作为区域表征的输入特征。

...

特征提取层

特征提取层从原始数据中抽取高维特征，用于支持后续表征学习。事实上是数据层的抽象，要不要合并为一个层？

- **图像与视觉特征**：使用CNN、Vision Transformer等提取RS和SV图像的嵌入。例：Tile2Vec

- **时空序列特征**：使用RNN、LSTM或TCN处理轨迹和流量数据，捕捉时序依赖。**例**：MTGRR、MobiCLR
- **图结构特征**：基于k-NN或相似性阈值构建邻接矩阵，如POI之间的余弦相似性。**例**：Multi-View Joint Graph Representation Learning for Urban Region Embedding
- **文本特征**：TF-IDF或LLM编码POI/新闻文本。**例**：UrbanCLIP

表征学习层

采用机器学习/图学习方法对区域进行向量化编码：

- **传统方法**：主成分分析（PCA）、自编码器（AE）、聚类、CNN。**例**：CGAP、SFRS
- **Transformer模型**：Attention机制等方法优化序列建模。**例**：ST-LLM、Terra、GURPP
- **图结构方法**：
 - 基于节点的图神经网络（GNN）。**例**：CGAP、FlexiReg
 - 基于邻接关系的图卷积网络（GCN）。**例**：HGI、RegionEncoder、DLCL
 - 结合多模态特征的图注意力网络（GAT）。**例**：Region2Vec、MVURE、MGRL4RE、USPM
 - 考虑时序演化的时空图网络（ST-GCN）。
 - 异质数据融合的异质图神经网络（Heterogeneous GNN）。**例**：HUGAT
- **LLM**：
 - 图生文：使用LLM，根据图像生成文本表征。**例**：CRL、UrbanCLIP、UrbanVLP
 - GeoLLM提取地理知识，CityGPT增强空间认知。

融合层

实现模态间的对齐与跨模态的协同。对提取到的不同维度、不同模态的表征向量进行融合：

- **简单拼接**：将不同模态的特征直接拼接在一起，形成高维特征向量。**例**：MP-VN、MVURE
- **三元组损失**：使用三元组损失进行嵌入学习，确保相似的区域被映射到相似的向量空间中。**例**：CaLLiPer、Urban2vec、RegionDCL
- **Attention Fusion**：通过注意力机制对不同特征进行加权融合，强调对任务最重要的特征。**例**：MGFN、MTGRR
- **对比学习**：通过对比学习优化区域表征，使得相似区域之间的表征更接近，不相似的区域表征之间距离更远。**例**：ReMVC、RegionDCL
- **LSTM**：融合时空特征。**例**：MMGR、USPM
- **视图间融合**：跨视图信息共享，互信息最大化和双向预测。**例**：ReCP

- 视图内融合：模态特定专家网络（MoE）后融合。例：MTGRR

任务层

针对不同的下游任务进行优化：

- Prompt 提示学习：使用LLM提示生成任务特定表征。GURPP
- LLM生成条件表征：使用对比语言-图像进行预训练。ReFound
- 强化学习：用于优化决策策略。

根据表征向量进行下游任务：

- 城市功能区划分：使用K-means进行土地使用分类。Tile2Vec、MVURE、REMVC
- 区域相似性分析：相同/不同城市区域间相似性分析。SFRS、MuseCL
- 城市动态预测：人流量、能耗、车流拥堵、碳排放、社会经济预测。MTGRR、UrbanCLIP、MuseCL
- 城市规划与模拟：犯罪率预测、区域流行度预测。HDGE、CGAP、MGFN、GURPP、FlexiReg

不同学习机制在范式中的角色

考虑到城市区域表征中使用了许多学习机制，这里做一个总结，看看后面怎么放。

学习机制	所处层级	作用机制	主要贡献
有监督学习	任务层	基于标签优化表征	提升特定任务性能
半监督学习	表征学习层+任务层	利用少量标签 + 自监督特征	提升低标注场景鲁棒性
自监督学习	表征学习层	无需标签，通过对比或重构任务学习特征	捕捉潜在功能一致性
对比学习	融合层+表征学习层	构建正负样本对以保持模态一致性	多模态对齐、跨城市迁移
强化学习	任务层	通过奖惩机制优化城市决策策略	城市规划、流量控制等生成式任务

范式总结

各层之间具有数据和信息的流动与传递关系。原始数据层提供多源异构的城市感知数据，特征提取层从这些原始数据中抽取高维特征，表征学习层基于提取的特征学习低维嵌入，融合层对来自不同模态的表征进行对齐与整合，任务层则使用融合后的最终区域表征向量来完成具体的下游任务。

早期范式主要依赖的单一模态，而现代范式强调多模态一致性和时空异质性，通过对比学习、LLM等增强泛化能力和跨城市能力。

存在的问题

1. **多模态异构性**：不同数据在空间、时间尺度和语义层次上差异很大，怎么有效融合这些异构信息？
2. **空间尺度差异与分区选择**：城市没有唯一空间划分方式，不同的任务可能采取不同的区域粒度。但是模型往往依赖固定的区域划分，在新的划分方式表现不佳
3. **语义漂移**：城市功能和社会经济特征会随时间动态变化，使得静态嵌入可能失效，需要设计时序更新和增量学习机制，此外还有数据不均匀问题，城市市区和郊区的数据密度有较大差异，其特征也十分不同，学习到的特征可能偏向数据丰富地区。
4. **可解释性与因果推理**：无法解释区域表征中隐含的语义，也尚未有系统方法将因果推理纳入城市区域表征中。

未来可能的发展方向：

1. 城市统一大模型

将不同模态、不同城市的知识统一预训练成大规模模型，让其可以作为所有城市下游任务的特征提取器，再使用Prompting或Fine-tuning来适应不同城市的不同下游任务。

2. 跨城市迁移与域自适应

不同paper中的区域划分方式不同，网格/行政区/道路单元/自适应聚类等等，如何实现不同划分方式的复用？不同城市之间的模型如何实现复用？

3. 可解释性与因果推理

当前学习到的表征一般都是高维向量，缺乏人类直观的理解。将因果图学习、解释性神经网络引入区域表征，探索“为什么”层面的空间语义机制。

4. 实时与动态城市表征

当前paper中的表征大多数都是静态的，如何捕捉现实中的动态信息并实时更新？实时交通流？怎样实现区域表征的时序更新与演化分析。

5. 数据公平性

整合影响表征公平性的数据，如残疾人设施分布等（周骁：城市老年人健康预测），让表征更公平、精准。

6. 时空融合

在城市区域表征中，最关键的两个模块是表征学习层和融合层，表征学习的效果直接影响嵌入向量对区域单一模态特征（如空间、语义或功能）的刻画深度，融合的效果直接影响最终表征能否全面、一致地整合多源异构信息，并决定其在下流任务中的泛化能力。这里重点对这两块进行扩展。

区域表征方法

城市区域表征方法旨在将城市各类多源数据映射为低维向量，以支持城市规划、功能区识别、人口预测等任务。现有方法可分为传统统计与特征工程、聚类与划分、深度表征学习等几大类，并强调多模态数据的整合与时空特征的捕捉。MTE、MGRL4RE

传统统计与特征方法

提取区域的统计特征，POI计数、类别分布、土地使用比例等，用于表达区域语义。这种方法提取的特征直观、易于解释，但难以捕获区域的复杂交互关系。CGAP对多源数据采用逐层粗化池化，输入依赖POI和人类流动。

聚类与分区方法

聚类方法用于区域划分和特征聚合，支持无监督表征学习：

- K-means或层次聚类：基于POI或轨迹相似性分区城市区域，用于土地使用分类或功能区识别。HGI
- 图分区：社区检测算法（Louvain）在路网或移动图上划分区域。
- 扩展：结合层次结构，使用POI的层次图来聚类语义相似区域。为后续嵌入提供粗粒度表示，但对噪声敏感。CGAP、HGI

表征学习方法

利用机器学习技术将区域嵌入向量化，捕捉隐含语义和时空特征。主要包括：

- 无监督学习与传统降维
 - PCA、自编码器等，用于提取主要特征，在标签稀缺时可提供粗略的区域表征。
- 时空序列模型
 - 采用RNN/LSTM或TCN等模型处理人流量、出租车轨迹等时序数据，捕捉出行时序模式。MobiCLR将区域日间的进出流量序列作为输入，通过对比学习捕获时序语义；
- 图神经网络：城市区域天然具有图结构特征，各类GNN被广泛用于城市区域表征。
 - GCN/GAT：RegionEncoder联合GCN与GAT学习卫星图像、POI、流量等多模态输入的区域表征
 - 异构图GNN：HUGAT构建异构城市图，将地理空间和人流变化纳入统一结构，通过元路径捕获不同语义，并采用HAN进行区域嵌入。
 - 时空图网络：ST-GCN融合时空信息，尝试时空图卷积来同时考虑空间拓扑和时间演化。
- 多图与多视图表征学习
 - MGRL4RE提出多图学习框架，通过构建包含POI、人类流动等多种关系的多图来融合区域间与区域内的关联，学习到既考虑空间语义又具备空间不变性的多视图表征。MTGRR使用多专家图结构，对点级（街景图）和聚合级（POI、人类流动）模态分别建模，并通过动态加权机制融合多模态。
- 对比学习
 - 采用对比损失优化区域嵌入，来拉近相似区域，拉远不同区域在表征空间中的距离。CGAP在融合层引入联合对比学习目标来提升嵌入质量。
- Transformer
 - 使用序列Transformer对长时序特征建模
 - 使用时空Transformer来处理轨迹数据
- LLM生成文本
 - 将图像或者图结构转换为文本描述，将非文本模态与LLM相结合，增强非文本模态的语义理解能力。CityGPT通过构建城市知识指令集和微调LLM强化对城市空间概念的理解，UrbanCLIP通过预训练大模型生成对卫星图像的文本描述，然后使用图文对比学习同一表征空间。

多模态融合策略

城市区域数据包括视觉、文本、图结构、时空序列等多种模态，因此融合不同模态信息至关重要。常用的多模态融合策略包括：

- 简单拼接
 - 将不同模态的特征向量直接拼接为高维向量。MVURE直接将POI、人流量等特征拼接后作为表征向量。方法简单，但无法确保各个模态的融合效果。
- 注意力加权融合
 - 使用注意力机制为每个模态赋予权重，动态聚合特征。MGFN在多图融合模块中使用了多层级交叉注意力，学习不同流动模式之间的信息交互，从而得到综合嵌入，MTGRR设计了空间自适应融合机制，根据区域位置动态调整各模态的融合权重。
- 对比学习+三元组损失
 - 构造正负样本，优化嵌入空间结构。Urban2Vec将街景图像和POI文本视为邻域文档，以文档嵌入方式联合编码不同模态信息。CaLLiPer提出对比语言-位置预训练框架，将POI文本与地理位置嵌入对齐，通过对比学习和三元组损失让语义相近的区域向量靠近。
- 融合层与多视图聚合
 - 通过专门的融合网络或层来融合不同模态间的特征。MGRL4RE构建多图来表示区域的不同关系，在表征学习模块中将各视图表征进行多视图融合。
- 文本-图像对比
 - UrbanCLIP将利用LLM生成卫星图像的文本描述与图像对齐，利用图像-文本对比预训练同一视觉和文本监督，从而增强区域视觉嵌入的语义信息，为图像特征注入了更细粒度的语义描述。
- 时空序列融合
 - 使用LSTM/TCN等对每一时序进行编码后再融合。部分模型会对不同时间窗口的人流序列提取特征，然后通过LSTM进行跨时间融合，从而获得动态的区域表征。

通过特征拼接、注意力加权、对比学习等机制，使得不同模态的信息可以在同一个表征空间中对齐并互补。这些策略促进了区域表征的泛化能力和跨城市、跨任务适用性。

论文	原始数据层	特征提取层	表征学习层	融合层	任务层
HDGE (CIKM'17)	人类流动数据 (如出租车行程记录)、空间区域相似性指标	构建流动图 (边权基于行程计数)和空间图(基于反距离权重)	通过图上随机游走生成序列,采用Word2Vec启发式嵌入学习	跨图联合随机游走实现统一嵌入	区域相似性度量、下游预测任务 (如犯罪预测、土地利用分类)

ZE-Mob (IJCAI'18)	出租车行程数据、城市区域划分、POI签到数据	提取人类流动模式（起源-目的地对、时间槽）、共现频率、PPMI矩阵、重力模型加权矩阵	PPMI矩阵因子分解（Word2Vec风格）、重力加权时空嵌入	在嵌入优化中整合共现和重力信息	功能区聚类、土地利用评估（NMI、ARI、F-measure、CID指标）
Measuring human perceptions (Landscape and Urban Planning'18)	街景图像、人类感知评分数据集	DCNN自动提取图像特征，分割为150个对象类别	DCNN用于预测人类感知指标（二分类任务）	RBF核SVM整合特征，多元回归融合视觉元素	人类感知预测（安全、活泼等6项指标）、城市感知映射、视觉元素相关性分析
Tile2Vec (AAAI'19)	卫星遥感图像切片	CNN提取图像补丁特征	基于空间邻近的对比学习	（单模态，无需多模态融合）	土地利用分类
MV-PN (AAAI'19)	POI数据、出租车GPS轨迹、签到数据	构建多视图POI-POI网络（地理距离、流动连通性）、区域间邻近关系（距离+功能相似度）	带非线性投影的自编码器，融入空间自相关性和top-k局部近似	多视图融合通过距离/流动视图损失求和，邻近融合采用top-k选择	区域流动流行度预测（SE、 R^2 、Tau、NDCG指标）
MVURE (IJCAI'20)	出租车轨迹数据、POI/签到属性、人口普查块	流动上下文（源/目标相关性）、属性相关性（TF-IDF、余弦相似度）	视图特定图注意力网络（GAT），跨视图共享联合学习	自注意力机制跨视图传播、自适应加权求和融合	土地利用分类（NMI、ARI）、犯罪预测（MAE、RMSE、 R^2 ）

Urban2Vec (AAAI'20)	街景图像、 POI文本数据	CNN提取图 像特征、TF- IDF/LLM编 码POI文本	多模态联合 嵌入学习 (扩展Skip- gram模型)	图像-文本对 比学习 (CLIP风 格)	城市规划分 析、人口特 征预测
Region Embedding with Intra and Inter- View Contrastive Learning (WWW'22)	POI数据、人 类流动数据	视图特定特征 提取 (POI类 别计数、流动 模式)	视图内对比 学习 + 视图 间对比学习	互信息最大 化对齐多视 图嵌入	土地利用分 类 (NMI、 ARI)、流行 度预测 (MAE、 RMSE、 R^2)
Beyond the First Law of Geography: Learning Representati ons of Satellite Imagery by Leveraging Point-of- Interests (WWW'22)	卫星影像、 POI数据	ResNet-18 CNN提取卫 星图像特征	POI视图对比 学习 (POI相 似对)、地 理视图对比 学习 (邻近 对)	注意力机制 融合POI视图 和地理视图 嵌入	社会经济指 标预测 (外 卖订单、评 论数、人 口、人口密 度) (R^2 、 RMSE、 MAPE)
Region2Vec (ArXiv'22)	出租车行程数 据 (起源-目 的地对)、地 理邻域关系、 POI属性	可达性/邻近/ 功能相关性 (余弦相似 度、知识图谱 上TransD嵌 入)	多图上 GAT, 带自 注意力/跨注 意力编码器- 解码器	全局编码器和视图特定 解码器实现 多图融合	区域聚类 (NMI、 ARI)、流行 度预测 (MAE、 RMSE、 R^2)

HUGAT (AAAI'22)	空间图（路网结构）、人类流动数据	构建异构城市图（节点：区域、POI类别、时间槽；边：空间/时间关系）	异构图注意力网络（HAN）结合元路径	元路径引导的多模态融合	犯罪率预测、收入估计、共享单车流量预测
MGFN (ArXiv'22)	多种流动数据图	多图融合模块提取流动模式	交叉注意力学习综合嵌入	跨模态流动模式融合	犯罪预测、签到预测、土地利用聚类
Urban Region Profiling With Spatio-Temporal Graph Neural Networks (T-ITS'22)	时空流动数据、区域属性特征	时空图结构提取	时空图神经网络学习区域剖面表示	时空信息融合	区域交通状态估计、剖面预测
Region-Wise Attentive Multi-View Representation Learning For Urban Region Embedding (CIKM'23)	多源城市数据（POI、轨迹、签到）	多视图特征提取（POI功能、空间邻近、流动模式）	区域感知多视图表示学习（ROMER框架）	区域感知注意力融合多视图嵌入	POI类别预测、人口预测（NMI、ARI、 R^2 ）

Geo-Tile2Vec (TSAS'23)	POI数据、轨迹数据、街景影像	Word2Vec风格动态流动事件提取	多阶段度量学习嵌入	早期流动嵌入与后期影像度量融合	POI分类、土地利用分类、餐厅价格/企业数回归预测
HGI (ISPRS'23)	POI数据（类别、位置信息）	类别嵌入、POI空间上下文图卷积	分层图池化（POI-区域-城市级别）、互信息最大化	多头注意力POI到区域聚合、区域图卷积	城市功能分布估计、人口密度预测、房价估计
Geographic mapping with unsupervised multi-modal representation learning from VHR image and vector data (ISPRS Journal'23)	VHR卫星图像、矢量数据（POI、道路）	CNN提取图像物理属性、one-hot编码矢量数据	多模态无监督对比学习	物理-语义多模态融合	地理映射、土地覆盖分类（OA、Kappa系数）
RegionDCL (KDD'23)	OSM建筑足迹数据、POI数据	CNN提取建筑形状特征、one-hot编码POI	双层对比学习（Transformer编码器）	建筑+POI特征拼接融合	土地利用分类、人口预测
ATGRL (ArXiv'23)	流动数据、POI数据、签到语义	多图构建整合流动/POI/签到信息	多图聚合捕捉局部和全局依赖	双阶段融合多视图表示（线性注意力）	城市规划分析下游任务

MVGCL (IEEE'23)	人类轨迹数据、空间邻接关系	轨迹和邻接图特征提取	多视图图对比学习	视图间对比对齐	区域表示下游任务
ReCP (AAAI'24)	POI类别数据、人类流动数据（进出流时序）	正/负样本增强、属性/流动视图编码器	视图内对比+重构损失、视图间对比+双向预测	互信息一致性最大化和熵最小化	土地利用聚类（NMI、ARI、F-measure）、流行度预测（MAE、RMSE、 R^2 ）
UrbanCLIP (WWW'24)	卫星图像 + 生成文本描述	图像编码、LLM生成文本描述	图像-文本对比预训练（CLIP模型）	融合自然语言和视觉语义	城市指标预测（ R^2 提升6.1%）
MuseCL (IJCAI'24)	遥感图像、街景图像、POI文本、人口流动数据	CNN（Inception v3/ResNet-18）提取图像特征、Skip-gram POI文本嵌入	多语义对比学习（Triplet Loss视觉、InfoNCE对齐）	注意力融合街景+遥感、POI语义对齐	社会经济指标预测（人口密度、住房密度、流动计数、POI数、评论数、犯罪）（ R^2 、RMSE）
CGAP (IJCAI'24)	POI数据、区域图、流动上下文	GNN处理区域图结构	分层图池化（Coarsened GNN）	融入POI与流动的多模态信息	可持续发展目标相关预测（SDG任务）
ReFound (KDD'24)	POI数据、卫星图像	预训练基础模型知识蒸馏提取POI/图像特征	基础模型再训练统一嵌入	语言-视觉基础模型蒸馏融合	城市区域理解、人口/房价预测（ R^2 、MAE）

USPM (KDD'24)	街景图像、空间邻接关系	预训练图像编码器提取街景特征	半监督预测模型（对比+邻接约束）	空间邻接图融合街景嵌入	城市街道剖析、社会经济映射 (R^2 、F1分数)
MTE (GIS'24)	人类轨迹数据（过渡序列、时间戳、位置）	过渡对、空间KNN图、时间比例向量	Skip-gram过渡嵌入、MVGRU空间/时间图学习	嵌入拼接、Voronoi面积加权聚合	土地利用分类、人口密度/房价预测
UrbanVLP (AAAI'25)	卫星影像、街景影像	CNN提取多层次视觉特征、LLM生成文本描述	多粒度视觉-语言预训练	宏观/微观信息融合、文本校准	社会经济指标预测（6项任务）
FlexiReg (KDD'25)	POI数据、土地利用标签、卫星/街景影像、文本描述	POI/土地利用/邻域GAT、ResNet提取影像特征	多模态网格嵌入、提示增强学习	自适应聚合融合、文本-区域/街景对齐	犯罪预测、签到预测、服务呼叫预测、人口预测
MGRL4RE (TIST'25)	POI数据、区域信息、人流数据	区域相关多图构建（空间语义不变特征）	多图表示学习	多视图嵌入融合	土地利用聚类、流行度/犯罪预测
AUAEC (Transportation Research Part C'25)	LBSN签到数据、出租车流数据、POI数据	路网空间插值、活动向量构建	视图间对比学习、活动感知对比	多视图对齐融合	土地利用分类、交通事故/延误/流量预测
CaLLiPer (Computers, Environment and Urban Systems'25)	POI位置数据、文本描述	位置编码、预训练文本编码器	多模态对比学习对齐位置-文本	对比对齐融合	土地利用分类、社会经济映射

MobiCLR (KAIST'25)	出行数据（出租车/网约车 时序流量）	LSTM处理进 出流序列	时序对比学 习与正则对 齐	对比损失拉 近进/出流嵌 入	收入预测、 教育水平 和社会脆弱 度估计
GURPP (KDD'25)	城市区域图 （POI、流 动、空间关 系）	图结构提取 （节点/边特 征）	子图中心预 训练模型	图提示融合 下游任务	区域表示下 游任务（土 地利用分 类、人口预 测）
EvolveURE(I nformation Fusion'25)	时空城市数据 （轨迹、POI 变化）	记忆模块捕捉 近期嵌入	记忆基图网 络生成演化 嵌入	时空记忆融 合	动态城市感 知、流量/事 件预测
MTGC (Information Fusion'26)	纽约/旧金山 交通流数据	构建不同时间 尺度的流量图	多尺度时序 图对比学习	不同时间图 对齐	犯罪预测、 签到预测、 土地利用聚 类